پروژه هشتم:

بررسی روش های مختلف آموزش ANFIS

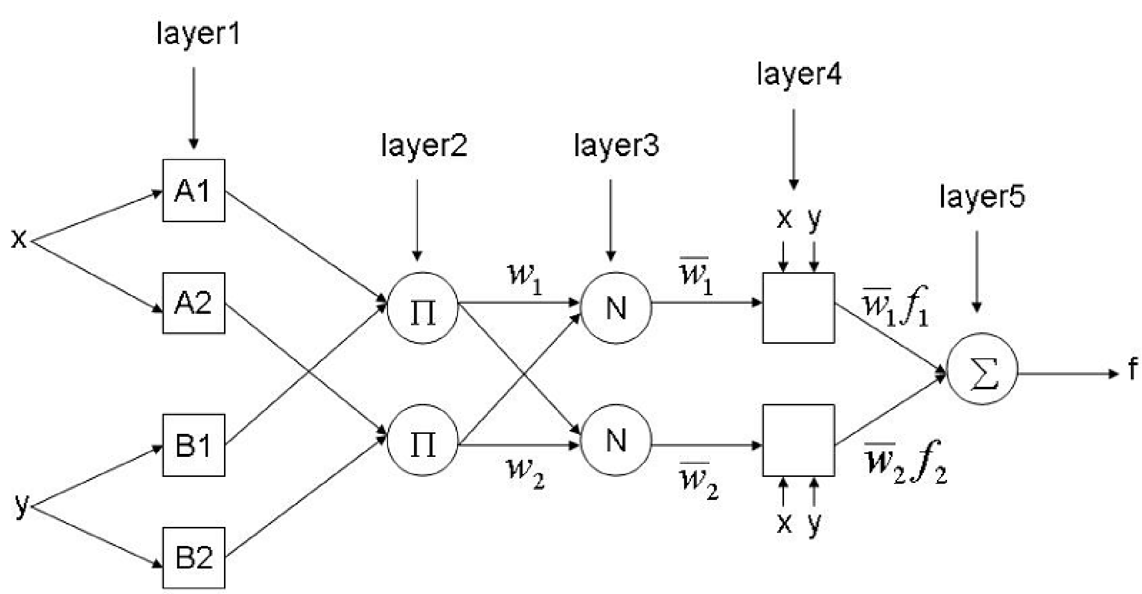
ANFIS(adaptive network-based fuzzy inference system) شبکه تطبیق پذیر و قابل آموزشی است که به لحاظ عملکرد کاملا مشابه سیستم استنتاج فازی است.
برای سادگی کار فرض می کنیم که سیستم فازی ما دو ورودی x و y دارد و خروجی آن z است. حال اگر قوانین به صورت زیر باشند :

$$Rule1: if\ x\ is\ A_1\ and\ y\ is\ B_1\ then\ f_1 = p_1 x + q_1 y + r_1$$
$$Rule2: if\ x\ is\ A_2\ and\ y\ is\ B_2\ then\ f_2 = p_2 x + q_2 y + r_2$$

و اگر برای غیر فازی ساز از غیر فازی ساز میانگین مراکز استفاده کنیم خروجی به صورت زیر خواهد بود:

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2} = \overline{w}_1 f_1 + \overline{w}_2 f_2 \quad st \quad \overline{w}_1 = \frac{w_1}{w_1 + w_2}, \ \overline{w}_2 = \frac{w_2}{w_1 + w_2}$$

ساختار معادل ANFIS به صورت زیر خواهد بود:



شکل ۱

خروجی لایه ها

لایه ۱: در این لایه ورودی ها از توابع عضویت عبور(membership functions) می کنند.

$$O_{1,i} = \mu A_i(x), \qquad for\, i = 1,2$$
$$O_{1,i} = \mu B_i(x), \qquad for\, i = 3,4$$

توابع عضویت هر تابع پارامتری مناسبی می تواند باشد که در اکثر موارد توابع گاوسین انتخاب می شوند. مثل تابع زنگی شکل عمومی:

$$\mu A(x) = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b_i}}$$

که $\{a_i, b_i, c_i\}$ مجموعه پارامترها هستند. پارامتر های این لایه به پارامتر های اولیه (premise parameters) معروف هستند.

لایه ۲: خروجی این لایه ضرب سیگنال های ورودی است که در واقع معادل قسمت اگر قوانین هستند.

$$O_{2,i} = w_i = \mu A_i(x) \mu B_i(y), \quad i = 1,2$$

لایه ۳: خروجی این لایه نرمالیزه شده لایه قبلی است:

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1,2$$

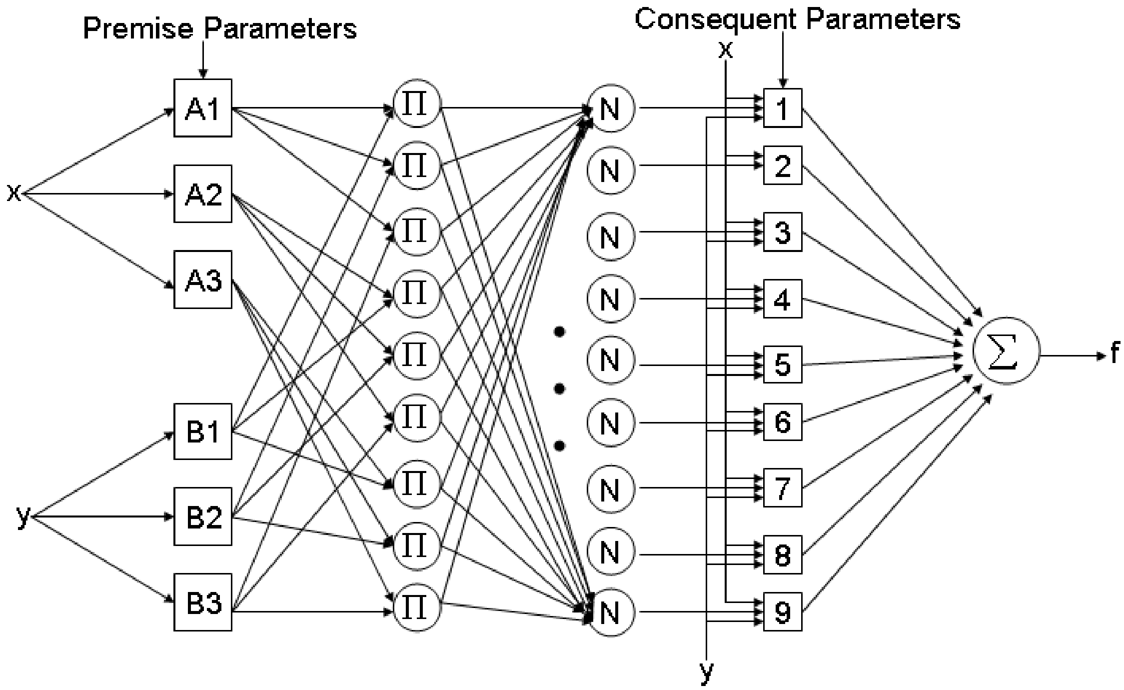
لایه ۴:

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i)$$

لایه ۵: خروجی این لایه خروجی کلی سیستم است:

$$O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

اکنون یک شبکه تولید شده است که معادل سیستم استنتاج فازی سوگنو است.
حال قرار است روش های آموزش چنین شبکه ای بررسی شود.
برای این کار ابتدا در لایه ۱ تمام قوانین موجود را تشکیل می دهیم.به طور مثال اگر ۲ ورودی داشته باشیم که هر کدام ۳ تابع عضویت داشته باشد ۹ قانون باید تشکیل دهیم.
که به صورت زیر خواهد بود.



شکل ۲

روشهای مختلف آموزش ANFIS

۱- Gradient Descent
۲- ترکیبی (hybrid)

۳- Levenberg-Marquardt

Gradient Descent

فرض می کنیم توابع تعلق ما به صورت گاوسین زیر باشند:

$$\mu A_i(x) = \exp(-(\frac{x-c_i}{a_i})^{2b_i})$$

معیار کارایی را هم مجموع مربعات خطا در نظر می گیریم.

$$E_k = \frac{1}{2}(y_k - o_k)^2$$

حال مشتق زنجیره برای پارامتر مثلا B_i به صورت زیر خواهد بود:

$$\frac{\partial E_k}{\partial B_i} = \frac{\partial E_k}{\partial o_k}\frac{\partial o_k}{\partial B_i} = -(y_k - o_k)\frac{\partial o_k}{\partial B_i}$$

که برای پارامتر های تالی (consequent parameters) به صورت زیر است:

$$\frac{\partial E}{\partial B_c} = \frac{\partial E}{\partial o_5}\frac{\partial o_5}{\partial o_4}\frac{\partial o_4}{\partial B_c}$$

و برای پارامتر های اولیه به صورت زیر است:

$$\frac{\partial E}{\partial B_p} = \frac{\partial E}{\partial o_5}\frac{\partial o_5}{\partial o_4}\frac{\partial o_4}{\partial o_3}\frac{\partial o_3}{\partial o_2}\frac{\partial o_2}{\partial o_1}\frac{\partial o_1}{\partial B_p}$$

حال مشتق لایه ها را نسبت به هم محاسبه می کنیم.

$$\frac{\partial o_5}{\partial o_4} = \frac{\partial(\sum f_j \overline{w}_j)}{\partial(f_i \overline{w}_i)} = 1$$

$$\frac{\partial o_4}{\partial o_3} = \frac{\partial(f_i \overline{w}_i)}{\partial(\overline{w}_i)} = f_i$$

برای مشتق لایه سوم به دوم اگر نود مقابل قانون متناظر را در نظر بگیریم به صورت زیر است:

$$\frac{\partial o_3}{\partial o_2}=\frac{\partial}{\partial w_i}\left(\frac{w_i}{\sum_{j=1}^{n} w_j}\right)=\frac{\sum_{j=1}^{n} w_j - w_i}{\left(\sum_{j=1}^{n} w_j\right)^2}$$

برای نود های غیر متناظر به صورت زیر است:

$$\frac{\partial o_3}{\partial o_2}=\frac{\partial}{\partial w_i}\left(\frac{w_j}{\sum_{j=1}^{n} w_j}\right)=\frac{-w_j}{\left(\sum_{j=1}^{n} w_j\right)^2}$$

مشتق لایه دوم نسبت به لایه اول به صورت زیر است:

$$\frac{\partial o_2}{\partial o_1}=\frac{\partial}{\partial A_m}\left(\prod_{A_j\in R(A_m)} A_j\right)=\prod_{A_j\in R(A_m),A_j\neq A_m} A_j$$

که در فرمول بالا $A_j \in R(A_m)$ مشخص کننده مجموعه های فازی است که بخش مقدمه قانونی را تشکیل می دهد که شامل مجموعه فازی A_m است.

در نتیجه مشتق مجموع مربعات خطا نسبت به پارامتر های تالی به صورت زیر است:

$$\frac{\partial E}{\partial B_c}=-(d-o)\frac{\partial o_4}{\partial B_c}$$

و برای پارامتر های اولیه به صورت زیر است:

برای نود های متناظر لایه سوم نسبت به دوم

$$\frac{\partial E}{\partial B_p}=-(d-o)f_i\frac{\sum_{j=1}^{n} w_j - w_i}{\left(\sum_{j=1}^{n'} w_j\right)^2}\prod_{A_j\in R(A_m),A_j\neq A_m} A_j\frac{\partial o_1}{\partial B_p}$$

برای نود های غیر متناظر لایه سوم نسبت به دوم

$$\frac{\partial E}{\partial B_p}=-(d-o)f_i\frac{-w_j}{\left(\sum_{j=1}^{n`}w_j\right)^2}\prod_{A_j\in R(A_m),A_j\neq A_m}A_j\frac{\partial o_1}{\partial B_p}$$

حال به بررسی مشتق لایه ۴ به پارامترهای تالی می پردازیم.

$$\frac{\partial o_4}{\partial p_i}=\frac{\partial}{\partial p_i}(f_i\overline{w}_i)=\frac{\partial}{\partial p_i}(\overline{w}_i(p_ix+q_iy+r_i))=\overline{w}_ix$$

$$\frac{\partial o_4}{\partial q_i}=\frac{\partial}{\partial q_i}(f_i\overline{w}_i)=\frac{\partial}{\partial q_i}(\overline{w}_i(p_ix+q_iy+r_i))=\overline{w}_iy$$

$$\frac{\partial o_4}{\partial r_i}=\frac{\partial}{\partial r_i}(f_i\overline{w}_i)=\frac{\partial}{\partial r_i}(\overline{w}_i(p_ix+q_iy+r_i))=\overline{w}_i$$

حال به بررسی مشتق لایه ۴ به پارامتر های اولیه می پردازیم.

$$\frac{\partial o_1}{\partial a_{ij}}=\frac{\partial}{\partial a_{ij}}\left(\exp\left(-\left(\frac{x-c_{ij}}{a_{ij}}\right)^{2b_{ij}}\right)\right)=2\frac{\left(\frac{x-c_{ij}}{a_{ij}}\right)^{2b_{ij}}b_{ij}\exp\left(-\left(\frac{x-c_{ij}}{a_{ij}}\right)^{2b_{ij}}\right)}{a_{ij}}$$

$$\frac{\partial o_1}{\partial b_{ij}}=\frac{\partial}{\partial b_{ij}}\left(\exp\left(-\left(\frac{x-c_{ij}}{a_{ij}}\right)^{2b_{ij}}\right)\right)=2\left(\frac{x-c_{ih}}{a_{ij}}\right)^{2b_{ij}}\ln\left(\frac{x-c_{ij}}{a_{ij}}\right)\exp\left(-\left(\frac{x-c_{ih}}{a_{ij}}\right)^{2b_{ij}}\right)$$

$$\frac{\partial o_1}{\partial c_{ij}}=\frac{\partial}{\partial c_{ij}}\left(\exp\left(-\left(\frac{x-c_{ij}}{a_{ij}}\right)^{2b_{ij}}\right)\right)=2\frac{\left(\frac{x-c_{ij}}{a_{ij}}\right)^{2b_{ij}}b_{ij}\exp\left(-\left(\frac{x-c_{ij}}{a_{ij}}\right)^{2b_{ij}}\right)}{x-c_{ij}}$$

در شبیه سازی های انجام شده از روش online backpropagation استفاده شده است

یعنی در هر تکرار به ازای هر داده که اعمال می شود خطا محاسبه می شود و پس انتشار خطا صورت می گیرد.

در ادامه نتایج شبیه سازی برای این روش مشاهده می شود.

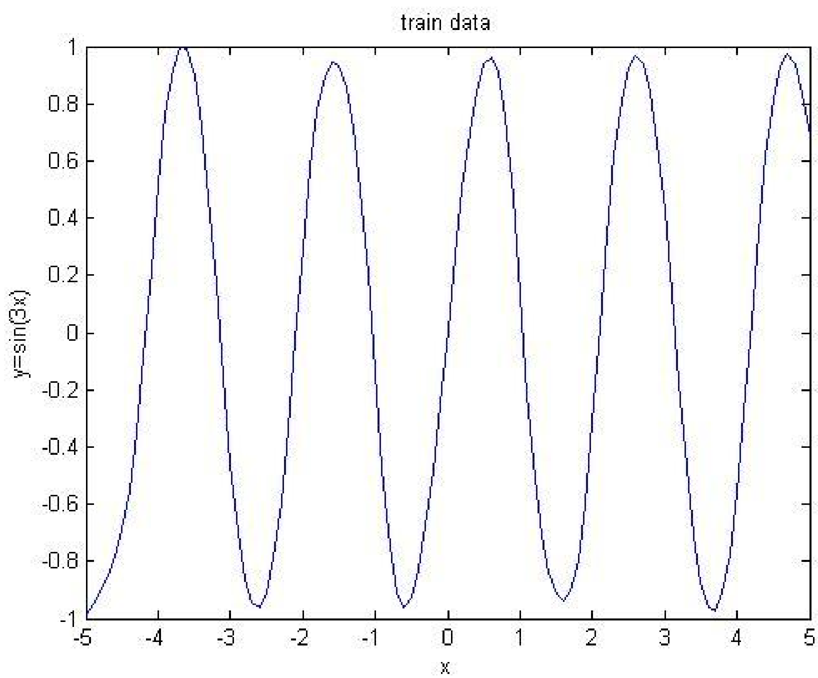
شبیه سازی

برای شبیه سازی از تابع زیر استفاده شده است:

$$y = \sin(3x), \qquad x \in [-5,5]$$

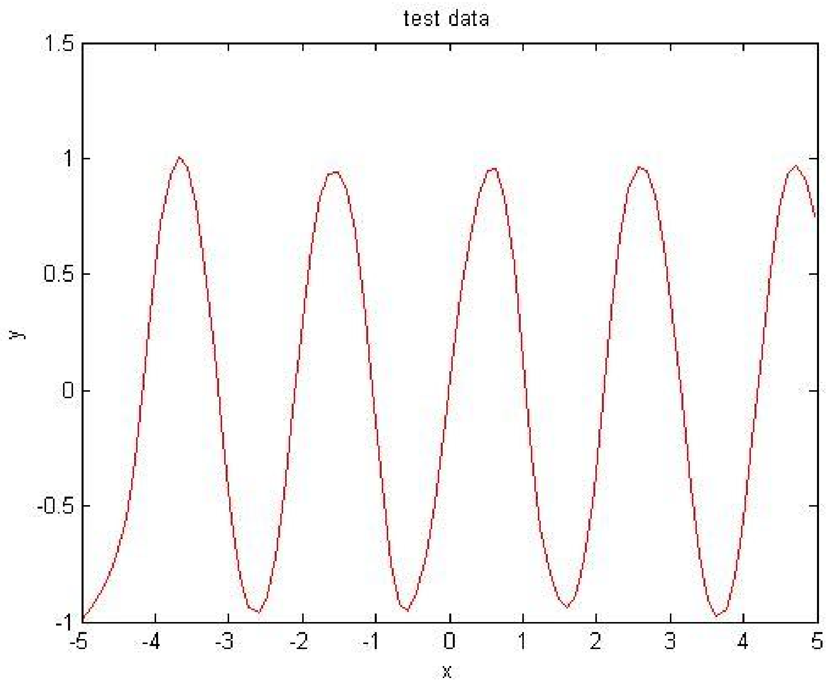
داده های آموزش را در این بازه با فاصله 0.1 از هم در نظر گرفته شده است و داده های تست با فاصله 0.12 .

ده بار شبیه سازی تکرار شده تا بتوان میانگین گرفت.



شکل ۳

شکل ۳ نتیجه به دست آمده توسط خود داده های آموزشی است.



شکل ۴

شکل ۴ مربوط به داده های تست است.

میانگین خطا برای داده های آموزش 0.3335 به دست آمده و برای داده های تست 0.2929 به دست آمده است.

روش ترکیبی(hybrid)

در این روش از ترکیب روش گرادیان نزولی و حداقل مربعات خطا (LSE) استفاده می شود. برای روشن شدن مطلب ابتدا روش LSE را مورد مطالعه قرار می دهیم.

حداقل مربعات خطا

در حالت کلی خروجی یک مدل خطی به صورت زیر است:

$$y = \theta_1 f_1(u) + \theta_2 f_2(u) + \ldots + \theta_n f_n(u)$$

که u بردار ورودی است و f ها توابع شناخته شده هستند و θ ها پارامتر های نامعلوم هستند که باید تخمین زده شوند. برای شناسایی پارامتر ها نیاز به داده های آموزشی داریم که به صورت زیر بیان می شوند.

$$\{(u_i; y_i), i = 1, \ldots, m\}$$

با جایگذینی زوج های ورودی و خروجی در معادله اصلی به m معادله خطی به صورت زیر می رسیم.

$$\begin{cases} \{f_1(u_1)\theta_1 + f_2(u_1)\theta_2 + \ldots + f_n(u_1)\theta_n = y_1 \\ \{f_1(u_2)\theta_1 + f_2(u_2)\theta_2 + \ldots + f_n(u_2)\theta_n = y_2 \\ \vdots \\ \{f_1(u_m)\theta_1 + f_2(u_m)\theta_2 + \ldots + f_n(u_m)\theta_n = y_m \end{cases}$$

که در فرم ماتریسی به صورت زیر است:

$$A\theta = Y$$

ماتریس A به فرم زیر است:

$$A = \begin{bmatrix} f_1(u_1) & \cdots & f_n(u_1) \\ \vdots & \vdots & \vdots \\ f_1(u_m) & \cdots & f_n(u_m) \end{bmatrix}$$

θ یک بردار $n \times 1$ است.

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$

Y بردار خروجی $m \times 1$ است.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

حال در این حالت بردار θ به صورت زیر قابل محاسبه است.

$\theta = A^{-1}Y$

البته تعدار جفت داده های ورودی خروجی معمولا از تعداد معادلات بیشتر است.

در حالت کلی داده ها ممکن است با نویز همراه باشند یا مدل ممکن است نتواند به طور دقیق خروجی را معین کند. در این حالت معادله به صورت زیر است.

$A\theta + e = Y$

در این حالت ما به دنبال $\theta = \hat{\theta}$ می گردیم که مجموع مربعات خطا را مینیمم کند.

$$E(\theta) = \sum_{i=1}^{m} (y_i - a_i^T\theta)^2 = e^T e = (y - A\theta)^T (y - A\theta)$$

اگر AA^T معکوس پذیر باشد آنگاه:

$$\hat{\theta} = (A^T A)^{-1} A^T y$$

روش دیگری نیز برای محاسبه پارامتر ها وجود دارد که یک روش بازگشتی است (RLSE) این روش به دو دلیل به وجود آمده است.

۱- گاهی ممکن است که ماتریس AA^T معکوس پذیر نباشد.

۲- فرض کنید θ را حساب کرده ایم اگر یک جفت داده ورودی خروجی جدید به سیستم اضافه شود در روش قبل دوباره باید کل پارامتر ها محاسبه شوند ولی در این روش فقط مقدار پارامتر جدید به دست می آید.

فرمول های بازگشتی را در زیر مشاهده می کنید.

$$S_{k+1} = S_k - \frac{S_k a_{k+1} a_{k+1}^T S_k}{1 + a_{k+1}^T S_k a_{k+1}}$$

$$\theta_{k+1} = \theta_k + S_{k+1} a_{k+1} (y_{k+1} - a_{k+1}^T \theta_k)$$

$$st \quad k = 0,1,\cdots m-1$$

برای شرایط اولیه $\theta_0 = 0$ و $S_0 = \gamma I$ در نظر گرفته می شود که I ماتریس همانی به ابعاد $m \times m$ است و γ یک عدد صحیح بزرگ است.

طریقه آموزش ANFIS با روش ترکیبی

در هر تکرار feedforward می رویم تا وقتی که ماتریس A که در روش LSE گفته شد به دست آید. خروجی ها را هم که داریم . سپس توسط روش ترکیبی پارامتر ها را به دست می آوریم. لازم به ذکر است که همه داده های آموزشی باید اعمال شوند و همچنین پارامتر های اولیه (premise patameters) ثابت نگه داشته می شوند. سپس پارامترهای تالی (conclusion parameters) ثابت نگه داشته می شوند و پارامترهای اولیه توسط گرادیان نزولی تنظیم می شوند.

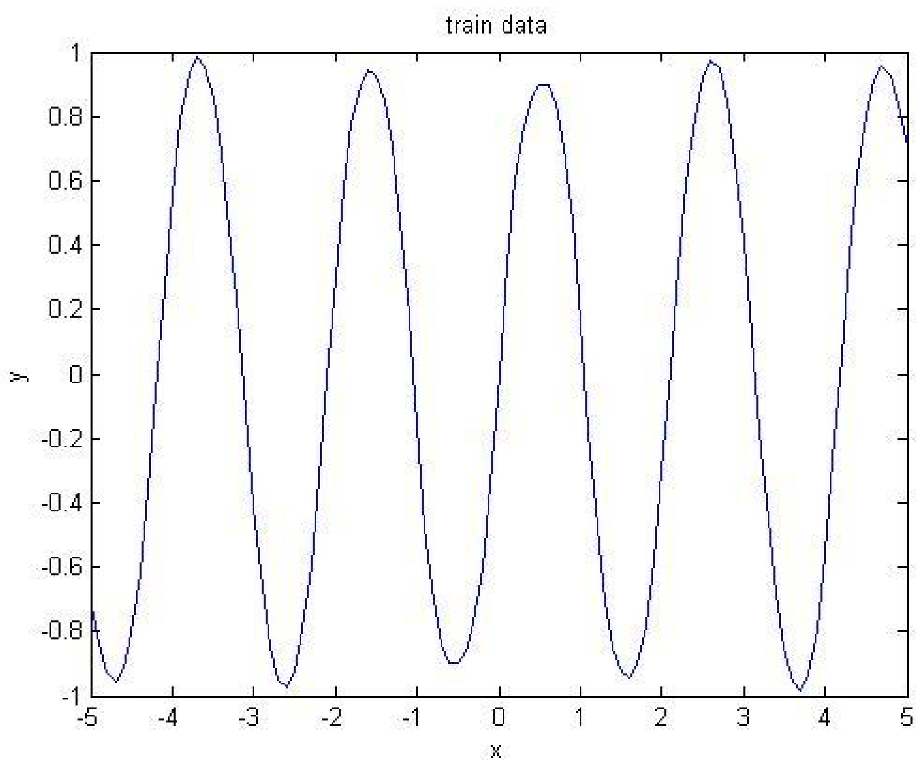
شبیه سازی

برای شبیه سازی از تابع زیر استفاده شده است:

$$y = \sin(3x), \qquad x \in [-5,5]$$

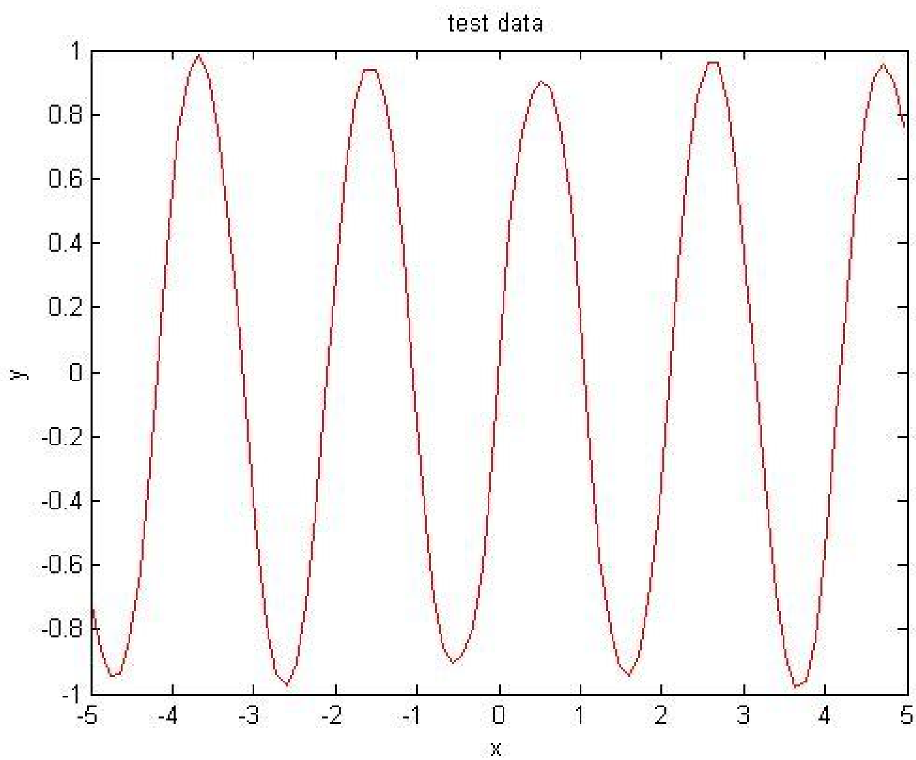
داده های آموزش را در این بازه با فاصله 0.1 از هم در نظر گرفته شده است و داده های تست با فاصله 0.12 .

ده بار شبیه سازی تکرار شده تا بتوان میانگین گرفت.



شکل ۵

شکل ۵ نتیجه به دست آمده توسط خود داده های آموزشی است.



شکل ۶

شکل ۶ مربوط به داده های تست است.

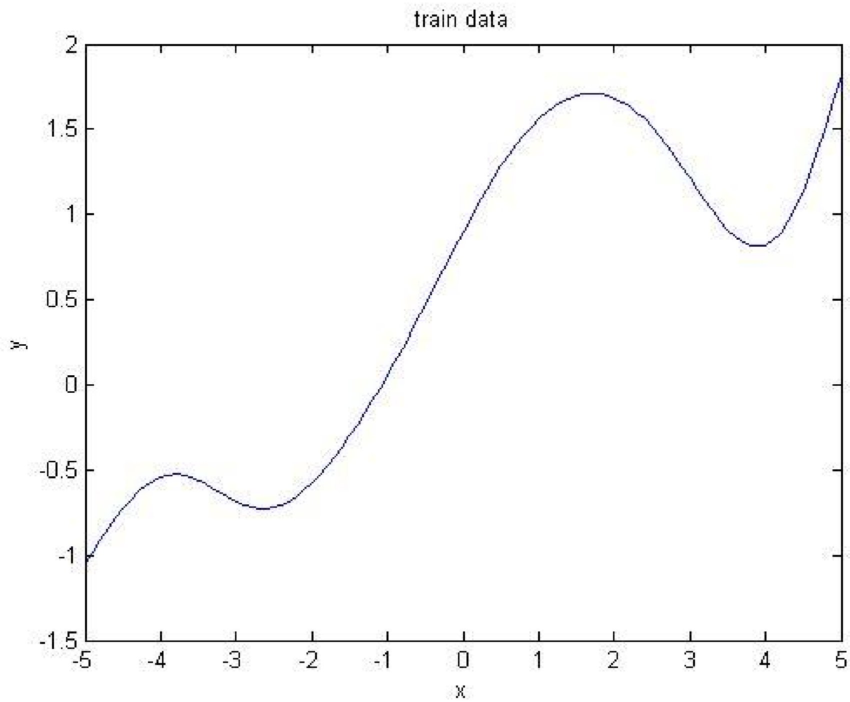
میانگین خطا برای داده های آموزش 0.1099 به دست آمده و برای داده های تست 0.1080 به دست آمده است.

مشاهده می شود که نسبت به روش گرادیان نزولی خطا کمتر است.

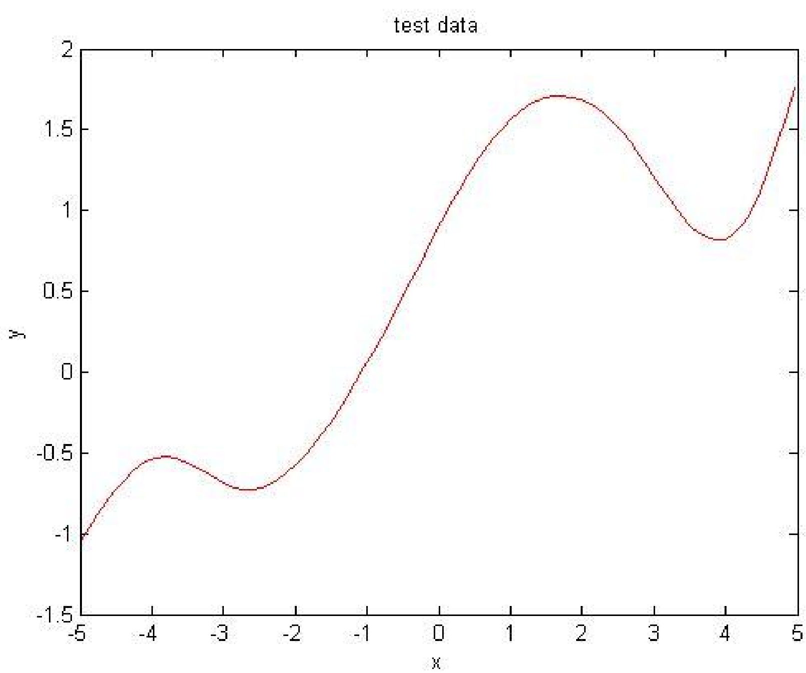
یکی از مزیت های اصلی روش ترکیبی سرعت آن است که بیشتر در کار online قابل استفاده است.

برای مشاهده این امر همان تابع بالا را در ۲ تکرار برای هر دو روش تکرار می کنیم و نتایج را مقایسه می کنیم.

ابتدا روش گرادیان نزولی:



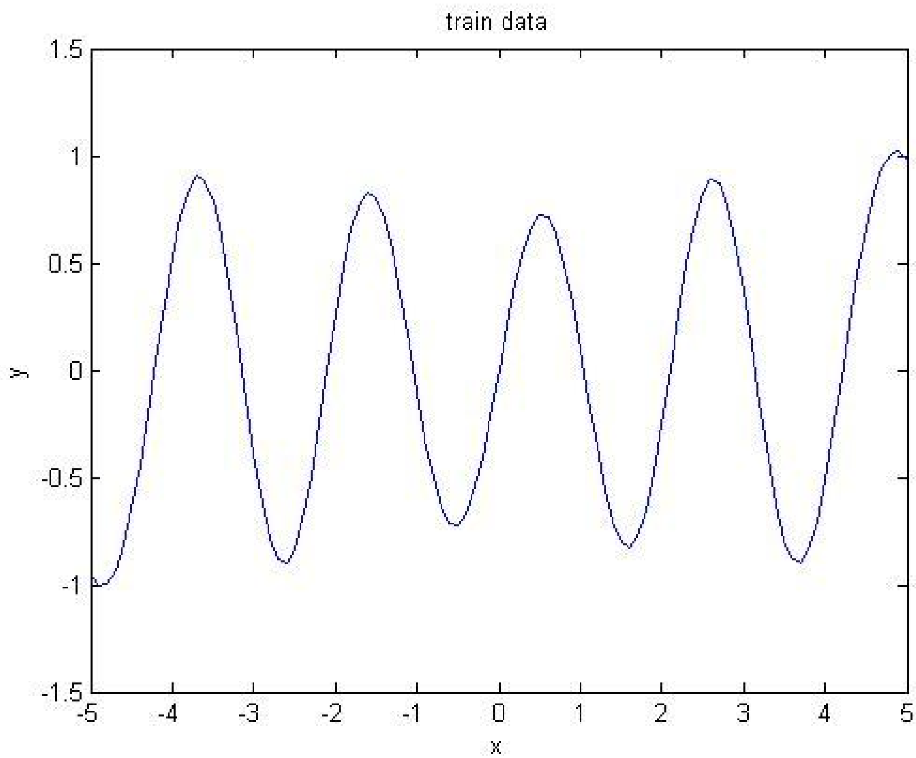
شکل ۷



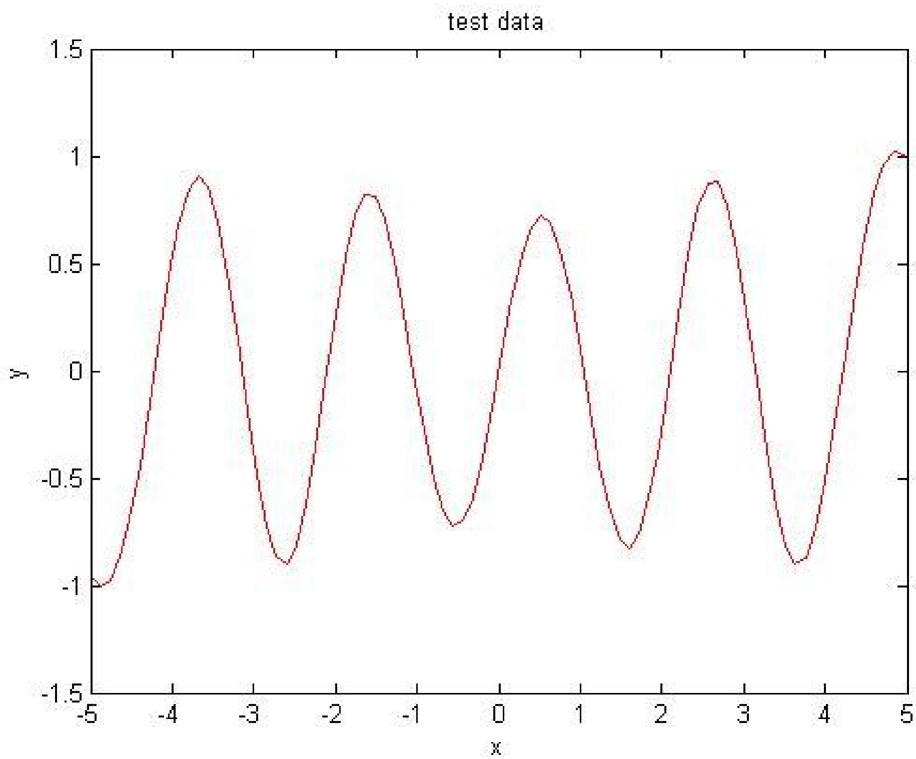
شکل ۸

میانگین خطا برای داده های آموزش 142.4810 به دست آمده و برای داده های تست 118.3182 به دست آمده است.
معلوم است که اصلا قادر به شناسایی تابع نیست.

روش ترکیبی:



شکل ۹



شکل ۱۰

میانگین خطا برای داده های آموزش 1.7152 به دست آمده و برای داده های تست 1.4018 به دست آمده است

مشاهده می شود که بسیار موفق تر از روش گرادیان نزولی عمل می کند.